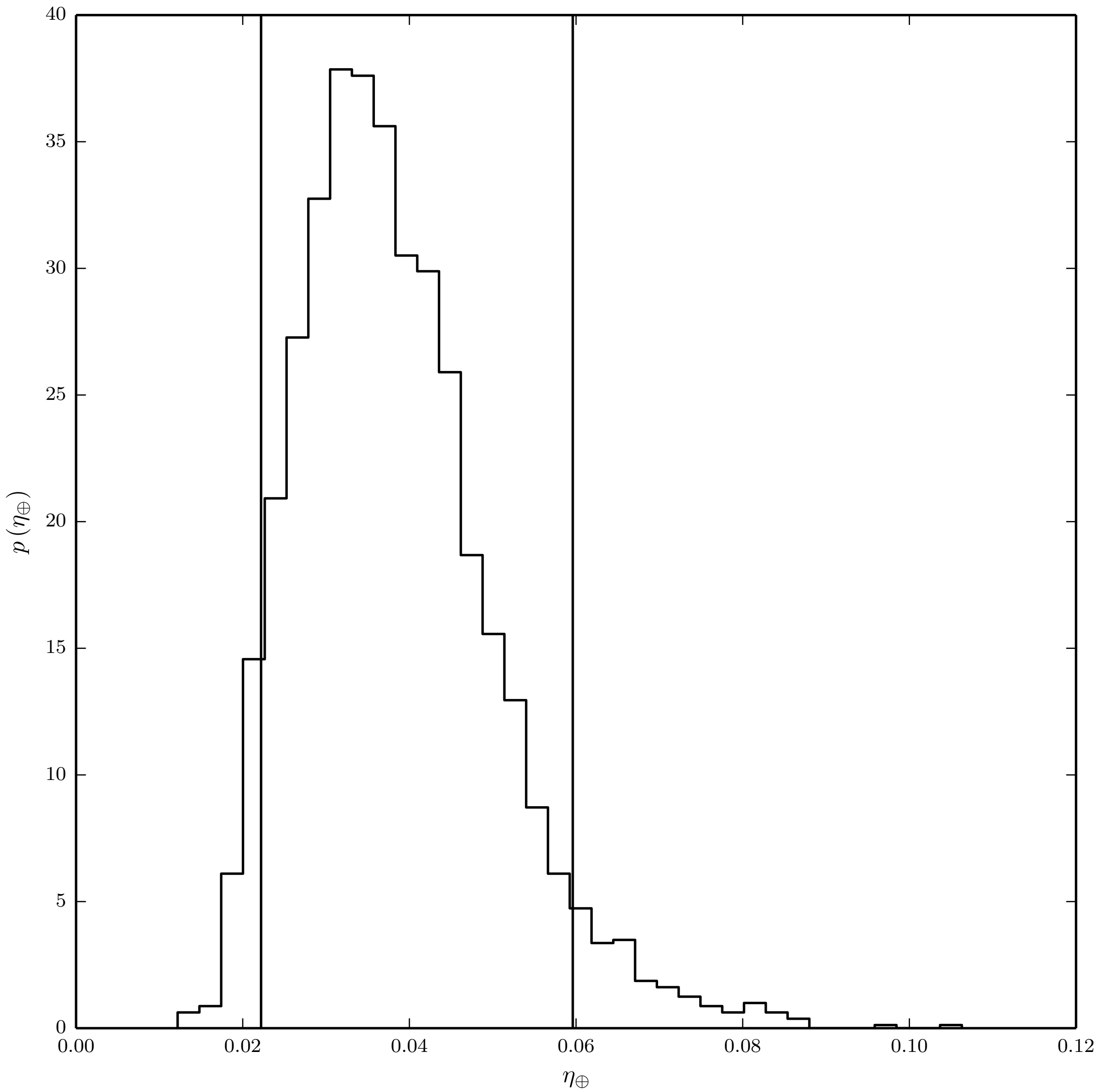
The Occurrence of Earth-Like Planets Around Other Stars

Will M. Farr[1], Chris Aldridge[1], Kirsty Stroud[1] & Ilya Mandel[1]

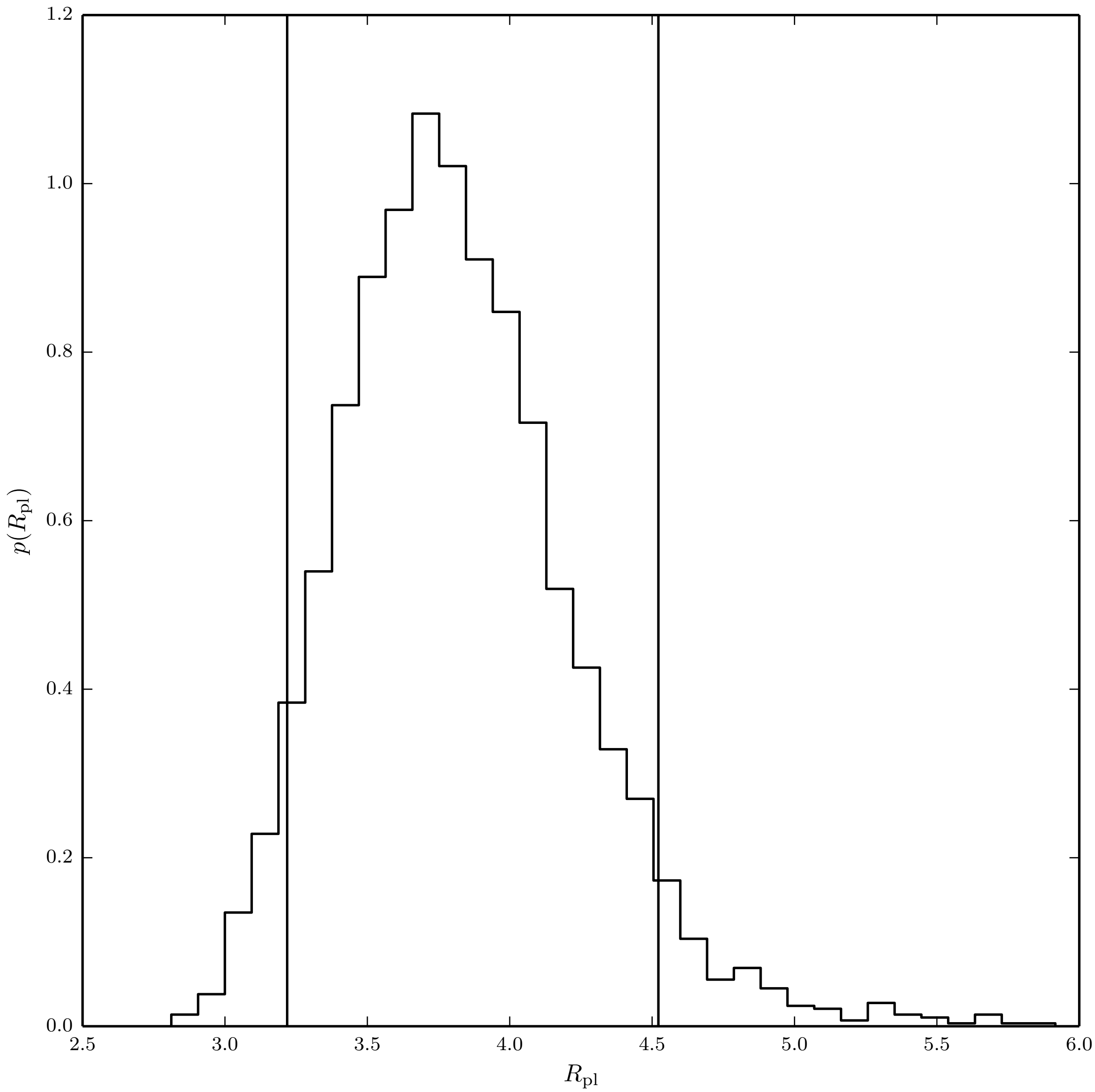
[1]*School of Physics and Astronomy, Univeristy of Birmingham, Birmingham, B15 2TT, United Kingdom*

The quantity $\eta_\oplus$, the number density of planets per star per logarithmic planetary radius per logarithmic orbital period at one Earth radius and one year periods, describes the occurrence of Earth-like extrasolar planets. Measurement of $\eta_\oplus$ is complicated by the difficulty of detecting Earth-like planets in Earth-like orbits about Sun-like stars. Previous estimates[1–4] place $1\% \lesssim \eta_\oplus \lesssim 34\%$. These works dealt with the problem of selection effects in the sample by either analyzing a region of the period-radius parameter space where observations are complete and extrapolating to $R = R_\oplus$ and $P = 1\mathrm{yr}$[1,2] or applying a binned analysis in the period-radius plane[3,4]. Here we present constraints on $\eta_\oplus$ from a parameterised forward model of the (correlated) period-radius distribution and the observational selection function in the most recent (Q17) data release from the Kepler satellite[5–7]. Our data set comprises 181,568 systems observed under the Kepler exoplanet observing program, producing 2598 planetary candidates. We parameterise the distribution of planetary periods and radii using a single, correlated Gaussian component; treat selection effects using a parameterised transit detection probability based on the measured noise level and stellar properties in the Kepler catalog; and include an empirically-parameterised, independent component in the planet period-radius distribution to represent false-positive planet detections. Using our model we

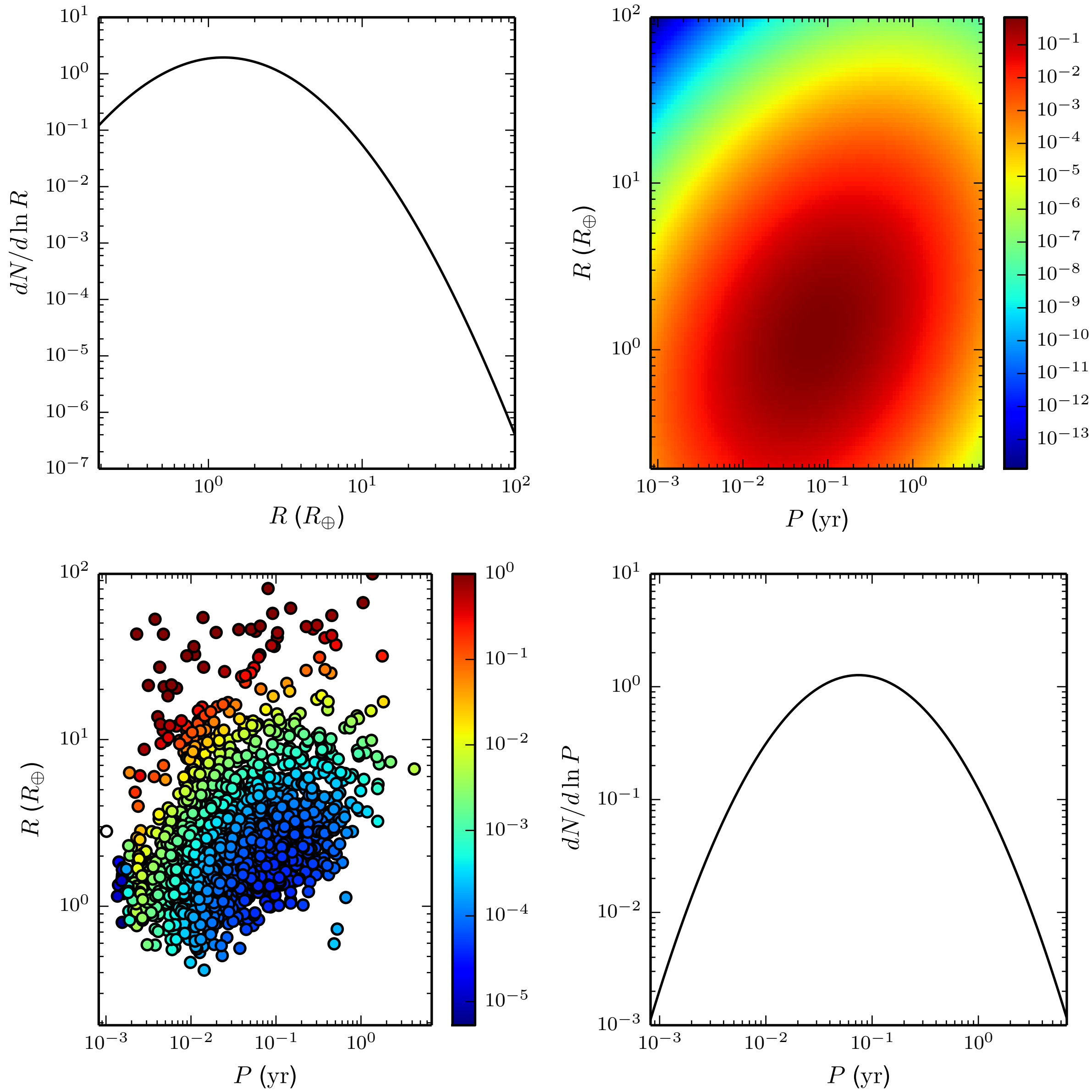
can simultaneously estimate $\eta_\oplus$, place constraints on the planet period-radius distribution function, and determine the degree of contamination by false-positive candidate indentifications. We find $\eta_\oplus = 3.8^{+2.1}_{-1.6}\%$ (90% CL). Additionally, we find that each star hosts $3.82^{+0.70}_{-0.60}$ planets with $P \lesssim 3\mathrm{yr}$ and $R \gtrsim 0.2 R_\oplus$, that the peak of the planet radius distribution lies at $R_{\mathrm{peak}} = 1.25^{+0.16}_{-0.17} R_\oplus$, and that $\ln P$ and $\ln R$ are correlated with correlation coefficient $r = 0.335^{+0.052}_{-0.054}$ (all 90% CL). Our empirical model for false-positive contamination is consistent with the dominant source being backgroud eclipsing binary stars[8], with $7.8^{+1.4}_{-1.3}\%$ (90% CL) of the candidates being false-positives.



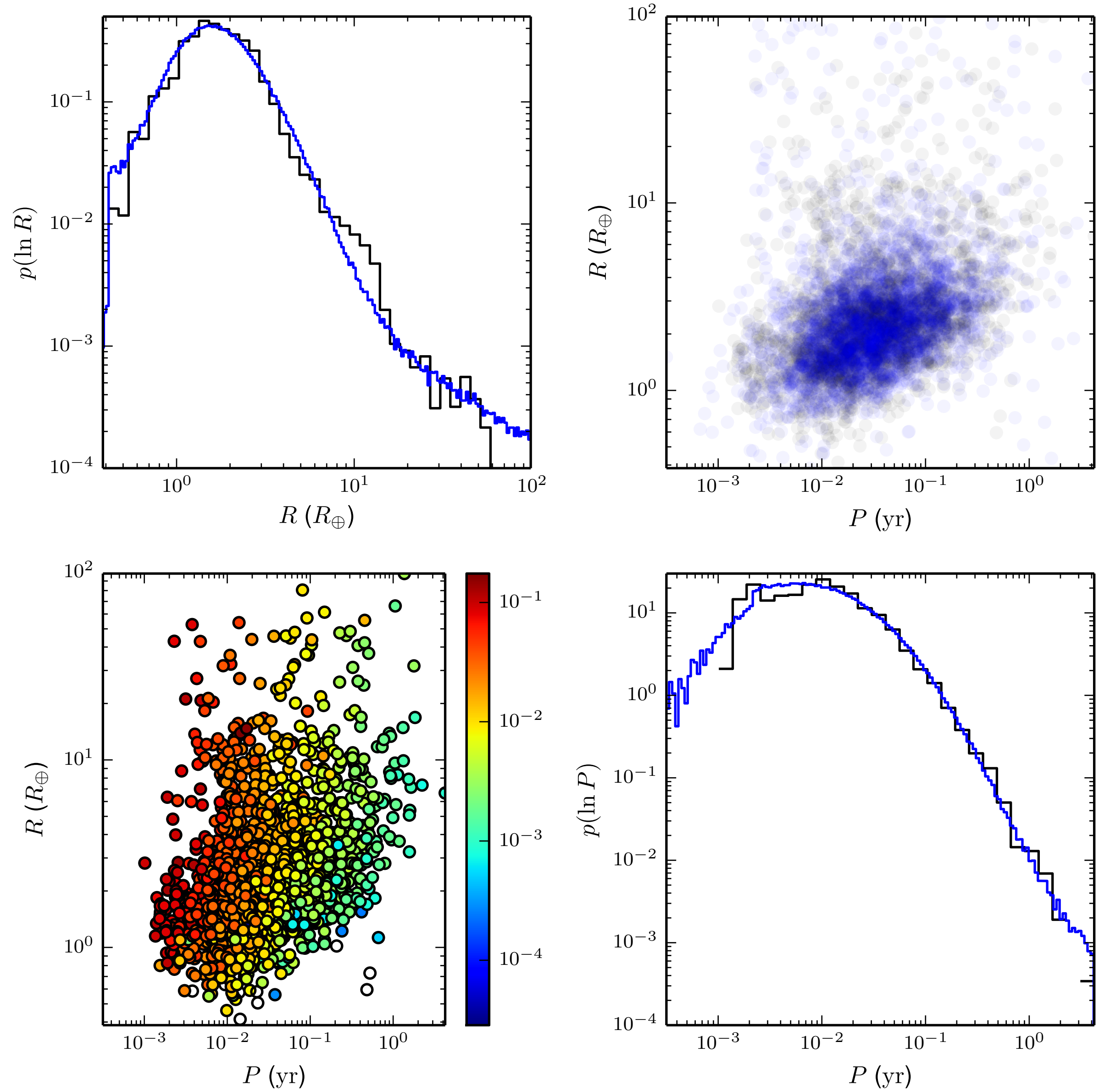
40
35
30
25
20
15
10
5
0
0.00
0.02
0.04
0.06
0.08
0.10
0.12
$p\,(\eta_\oplus)$
$\eta_\oplus$



1.2
1.0
0.8
0.6
0.4
0.2
0.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
$p(R_{\rm pl})$
$R_{\rm pl}$



dN/d ln R
R (R⊕)
R (R⊕)
P (yr)
R (R⊕)
P (yr)
dN/d ln P
P (yr)



1. Catanzarite, J. & Shao, M. The Occurrence Rate of Earth Analog Planets Orbiting Sun-like Stars. *The Astrophysical Journal* **738**, 151 (2011). `arXiv:1103.1443`.

2. Traub, W. A. Terrestrial, Habitable-zone Exoplanet Frequency from Kepler. *The Astrophysical*

Journal **745**, 20 (2012). arXiv:1109.4682.

3. Dong, S. & Zhu, Z. Fast Rise of ”Neptune-size” Planets (4-8 R $_{\oplus}$) from P ˜ 10 to ˜250 Days—Statistics of Kepler Planet Candidates up to ˜0.75 AU. *The Astrophysical Journal* **778**, 53 (2013). arXiv:1212.4853.

4. Petigura, E. A., Howard, A. W. & Marcy, G. W. Prevalence of Earth-size planets orbiting Sun-like stars. *Proceedings of the National Academy of Science* **110**, 19273–19278 (2013). arXiv:1311.6806.

5. Borucki, W. J. *et al.* Kepler Planet-Detection Mission: Introduction and First Results. *Science* **327**, 977– (2010).

6. Borucki, W. J. *et al.* Characteristics of Planetary Candidates Observed by Kepler. II. Analysis of the First Four Months of Data. *The Astrophysical Journal* **736**, 19 (2011). arXiv:1102.0541.

7. Batalha, N. M. *et al.* Planetary Candidates Observed by Kepler. III. Analysis of the First 16 Months of Data. *The Astrophysical Journal Supplement* **204**, 24 (2013). arXiv:1202.5852.

8. Fressin, F. *et al.* The False Positive Rate of Kepler and the Occurrence of Planets. *The Astrophysical Journal* **766**, 81 (2013). arXiv:1301.0842.

Acknowledgements WMF and IM are supported by a STFC consolidated grant number NNNN. Computations in this work were performed on the University of Birmingham's BlueBEAR cluster. Some/all of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST).

STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX13AC07G and by other grants and contracts. This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission directorate.

Author Contributions All authors assisted in the computational modelling, discussed the results, and edited the manuscript.

Reprints Reprints and permissions information is available at `www.nature.com/reprints`.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to W.M.F. (email: `w.farr@bham.ac.uk`).

Figure 1 The posterior on $\eta_\oplus$ accounting for selection effects and false positives. Recall that $\eta_\oplus \equiv \left.\frac{dN_{\rm pl}}{d\ln P\, d\ln R}\right|_{P=1{\rm yr}, R=1R_\oplus}$. Vertical lines indicate the 90% credible range. We find $\eta_\oplus = 3.8^{+2.1}_{-1.6}\%$, slightly lower than but consistent with previous work[4].

Figure 2 The posterior on $R_{\rm pl}$. Recall that $R_{\rm pl}$ is the number of planets per star. Vertical lines indicate the 90% credible range. We find $R_{\rm pl} = 3.82^{+0.70}_{-0.60}$. Will: CITE/discuss previous work.

Figure 3 The inferred planet period-radius distribution accounting for selection effects and false-positives. (Upper Left) The planet number density per logarithmic planet radius. The density peaks at $R_{\rm peak} = 1.19^{+0.17}_{-0.18} R_\oplus$ (90% CL). (Upper Right) The planet number density in the period-radius plane. The inferred correlation coefficient between $\ln P$ and $\ln R$ is $r = 0.335^{+0.052}_{-0.054}$. (Lower Left) Scatter plot of the radius and period of the Kepler planet candidates. Color indicates the posterior false-positive probability for each candidate. Overall, the model prefers a false-positive rate of $7.8^{+1.4}_{-1.3}\%$ (90% CL). The primary contaminant is probably background eclipsing binaries; our rate is consistent with previous work[8]. (Lower Right) The planet number density per logarithmic planet period. The density peaks at $P = 0.075^{+0.007}_{-0.006}{\rm yr}$.

Figure 4 Comparison of synthetic data sets produced from the forward model incorporating selection effects with observed candidates. (Upper Left) The observed (black curve) and synthetic (blue curve; including planets and false positives, and using the fitted selection model to down-select the candidates from the planet distribution) normalised candidate density per logarithmic radius. Except for a discrepancy at $R \simeq 10 R_{\oplus}$—associated with hot Jupiters, a distinct planetary populationWill: CITE, and check—the model produces a good fit to the observed candidates over the range of reported radii. (Upper Right) Scatter plot of the observed candidates (black circles) and a posterior-averaged draw of candidates from the model (blue circles). (Lower Left) Scatter plot of the observed candidates. Colors indicate the posterior-averaged selection probability for each planet about its host star. Recall that the selection probability is treated a product of a geometric factor giving the probability of an isotropically-oriented orbit producing a transit and a signal-to-noise-ratio-dependent transit detection probability. (Lower Right) The observed (black curve) and synthetic (blue curve; including planets and false positives, and using the fitted selection model to down-select the candidates from the planet distribution) normalised candidate density per logarithmic period. Except for the aforementioned hot Jupiter peak at $P \simeq 1 \mathrm{day}$ the model produces a good fit to the observed candidates over the range of reported periods.